N

negative binomial distribution– For a series of **Bernoulli trials,** the negative binomial distribution gives the **probability** of the total number of **trials** (n) required to obtain k successes. It is given by the formula

$$P(n) = \binom{n-1}{k-1} p^k (1-p)^{n-k} \qquad n = k, k+1, \ldots$$

where p is the probability of success on a single trial. It is also called the Pascal distribution. Note that the **geometric distribution** is a special case of the negative binomial distribution with $k = 1$.

negative correlation– In **correlation analysis,** two **variables** are said to have negative correlation when high values of one variable tend to be associated with low values of the other and vice versa. Some examples of negative correlations are selling price and demand, absenteeism and production output, sales and competitors' expenditure on advertising, among others. The concept applies only to pairs of variables, i.e., to **simple correlation.** It does not apply to **multiple correlation.** See also *coefficient of correlation, inverse relationship, positive correlation.*

negatively skewed distribution– See *skewed distribution.*

negative multinomial distribution– A generalization of the **negative binomial distribution** to the **sampling** involving a **multinomial experiment.**

negative predictive value– Same as *predictive value negative.*

negative relation– Same as *negative correlation.*

negative relationship– Same as *negative correlation.*

negative skewness– See *skewed distribution.*

negative study– A study that fails to establish the viability of the **research hypothesis.** A negative study does not result in the rejection of the **null hypothesis** and the results are **statistically nonsignificant.**

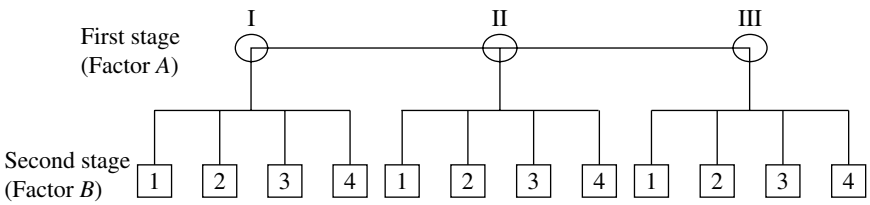
negative synergism– See *synergism.*

neonatal death rate– See *infant death rate.*

neonatal mortality rate– Same as *neonatal death rate.*

nested case-control study– A type of **case-control study** in which a **cohort** is followed through a period of time to select **cases** of interest, and for each case the **controls** are selected from within the cohort.

nested design– An **experimental design** in which **levels** of one or more **factors** are nested within one or more other factors. More specifically, given two factors *A* and *B*, the levels of *B* are said to be nested within the levels of *A* if each level of *B* appears with only a single level of *A* in the **observations.** For example, an **experiment** may be designed where water samples are taken from different sources of water supply. Here, water samples are nested within sources of water supplies. Similarly, in a simple **parallel group design,** patients receive only one **treatment,** i.e., patients are nested within treatments. Such designs are common in many fields of study and are particularly popular in **surveys** and industrial experiments. Compare *factorial design.*



A layout for the two-way nested design

nested model– An **analysis of variance** model involving a **nested design.** See also *crossed model, crossed-nested model.*

nesting– A term sometimes used to describe the characteristic or propensity of a **nested design.**

net reproduction rate– A measure of the rate of replacement of females in the population per generation, with the current values of **fertility** and **mortality.** In a **cohort** subject to a given set of **age-specific fertility rates, age-specific mortality rates,** and given sex **ratio** at birth, it is the **average** number of female children born per woman. See also *gross reproduction rate.*

Newman–Keuls test– A type of **multiple comparison** procedure for comparing pairwise **means** following a significant ***F* test** in an **analysis of variance.** The procedure involves a step-by-step approach where the **sample ranges** are tested against the **Studentized range** of the subsets rather than the range of the mean values. See also *Bonferroni procedure, Duncan multiple range test, Dunnett's multiple comparison test, Scheffe's test, Tukey's test.*

Newton–Raphson method– A numerical **algorithm** normally employed for optimization of a mathematical function. The procedure involves solving equations iteratively, in which each successive approximation is determined by using the first derivative of its numerical estimates.

nominal category– A category or group defined by a **nominal** or **categorical variable.**

nominal data– Data obtained by using nominal **scales of measurement.** See also *categorical data, nominal scale, numerical data, qualitative data.*

nominal level of measurement– Same as *nominal scale.*

nominal measure– Same as *nominal variable.*

nominal scale– Measurement scales representing qualitative differences among categories or groups. Numbers may be assigned for purposes of identification, but the measure assigned to an item is simply a label used for identification. A nominal scale has only one property, class inclusion/exclusion for each one of the categories or classes; no quantitative relationships between classes are referred to or implied. Nominal scales produce **nominal** or **categorical data.**

nominal significance level– A term used to denote the actual **level of significance** of a **statistical test** when all its **assumptions** are satisfied.

nominal variable– Same as *categorical variable.*

nomogram– A **graphical representation** for the **variables** involved in a formula on a plane surface. It shows scales for the variables, their relative magnitudes, and their positions in manners such that the corresponding values of the variables are found at the points on the scales that are intersected by the same straight line. It is also called nomograph and alignment chart.

nomograph– Same as *nomogram.*

nomographic– Pertaining to the **graphical device** used in a **nomograph.**

nonadditive model– A **statistical model** in which the **explanatory variables** do not have an additive effect on the **response measure** of interest. In a **factorial experiment,** the term is used to refer to the tendency for the combination of **factors** to yield a result that is different from the sum of their individual contributions. Compare *additive model.* See also *interaction.*

nonadditivity– See *Tukey's test for nonadditivity.*

nonbalanced data– Same as *nonorthogonal data.*

noncentral chi-square distribution– See *noncentral distributions.*

noncentral distributions– The term is applied to a number of **probability distributions** that are closely linked to some commonly used **sampling distributions,** such as t, χ^2, and F distributions, and that arise in the form of the distributions of the **test statistics** derived under some specified **alternative hypotheses.** Some of the well-known distributions are noncentral t, χ^2, and F distributions. These distributions are useful in calculating the power of the tests on the basis of the corresponding central distributions. Some further details on noncentral t, χ^2, and F distributions are given in App. E.

noncentral *F* distribution– See *noncentral distributions.*

noncentral *t* distribution– See *noncentral distributions.*

noncompliance– A term used to describe the behavior of patients who do not follow one or more of the guidelines laid in the study **protocol.**

nondirectional hypothesis– An **alternative hypothesis** that does not indicate the direction of the possible differences from the value specified by the **null hypothesis.** See also *directional hypothesis, one-sided hypothesis.*

nondirectional test– Same as *two-tailed test.*

nonindependent events– In **probability theory,** when the occurrence of one **event** influences the **probability** of occurrence of another event, such events are said to be nonindependent. See also *dependent events, independent events.*

nonindependent samples– Same as *dependent samples.*

noninformative prior– A term used in **Bayesian statistics** to describe a prior whose **probability distribution** does not contain any empirical or theoretical information regarding the unknown **parameters.** An example of a noninformative prior is a **uniform distribution.** It is alternatively known as **diffuse** or **vague prior.**

nonlinear model– A **statistical model** in which the **parameters** are nonlinear. For example, the model $y = \alpha e^{-\beta x} + \varepsilon$ represents a nonlinear model. Some of the nonlinear models can be converted into **linear models** by making appropriate mathematical **transformations.** See also *nonlinear regression.*

nonlinear regression– The method of determining a **regression** in which a curve other than a straight line best describes the relationship between two **variables.** This is also called curvilinear regression. A nonlinear regression is based on a **nonlinear model.** Compare *linear regression.*

nonlinear regression model– A **regression model** that is nonlinear in the **parameters.** See also *nonlinear model, nonlinear regression.*

nonmutually exclusive events– In **probability theory,** two or more **events** are said to be nonmutually exclusive if the occurrence of one event does not preclude the occurrence of the other events. Compare *mutually exclusive events.*

non-normal distribution– A **probability distribution** other than a **normal distribution.**

non-normality– A term used to denote the property of a **random variable** having a **non-normal distribution.**

non-normal probability distribution– Same as *non-normal distribution.*

nonorthogonal data– Experimental data obtained by using a **nonorthogonal design.** Compare *orthogonal data.*

nonorthogonal design– A term used to denote an **analysis of variance** design with two or more **factors** having unequal numbers of **observations** in each **cell.** Compare *orthogonal design.*

nonparametric analysis– See *nonparametric methods.*

nonparametric methods– Methods of testing a **hypothesis** or obtaining a **confidence interval** that do not require knowledge of the form of the underlying **parent population.** These are statistical methods or tests that do not involve the **estimation** or **hypothesis testing** of **population parameters.** They are also called **distribution-free methods,** since they supposedly do not require that the underlying distributions be either normal in **shape** or homogeneous in terms of **variance.** The **data** that exhibit **positive** or **negative skewness**

can be analyzed by nonparametric methods. These methods can be applied when only **rank order** or preference data are available. In many cases, these methods are only slightly less powerful than their parametric analogues that assume a specific form of the population distribution (usually a **normal distribution**), even when that assumption is true. The nonparametric methods include the **Mann–Whitney *U* test; Wilcoxon signed-rank** and **rank-sum tests; Kruskal–Wallis** and **Freidman tests; Pearson chi-square test;** and the **Spearman rank correlation, point biserial, phi,** and **Cramér's *V* coefficients,** among others.

nonparametric procedure– Same as *nonparametric method.*

nonparametric regression– A **regression model** that does not assume any parametric form. There are currently a number of techniques for performing a nonparametric regression.

nonparametric statistical methods– Same as *nonparametric methods.*

nonparametric statistical test– See *nonparametric methods.*

nonparametric techniques– Same as *nonparametric methods.*

nonparametric test– Same as *nonparametric statistical test.*

nonprobability sample– A **sample** selected in such a manner that the **probability** of each element being selected in the sample is unknown. **Convenience** and **judgment samples** are examples of nonprobability samples. See also *probability sample, random sample.*

nonprobability sampling– Any **sampling** procedure in which the **probability** of an element being included in the **sample** is not known.

nonrandomized clinical trial– A **clinical trial** in which patients are assigned to **treatment** and **control groups** by some subjective criteria or a mechanism other than a randomized procedure. Such a trial is subject to several sources of **biases.** For example, patients who respond to a treatment may be healthier than those who do not respond, giving a false impression that the treatment is beneficial.

nonrandom sampling– Same as *nonprobability sampling.*

nonrecursive model– A **causal model** in which there is two-way causal flow in the system. Compare *recursive model.* See also *path analysis, structural equation model.*

nonresponse– A term used to denote the lack of response on the part of the respondent or the failure to obtain the relevant information being collected in a **survey.** The general problem of nonresponse arises because the characteristics of nonrespondents usually differ to some degree from those of respondents. A nonresponse can occur for a number of reasons (such as absence, death, or refusal to reply) and a high **nonresponse rate** can introduce **bias** into the results. See also *nonresponse bias.*

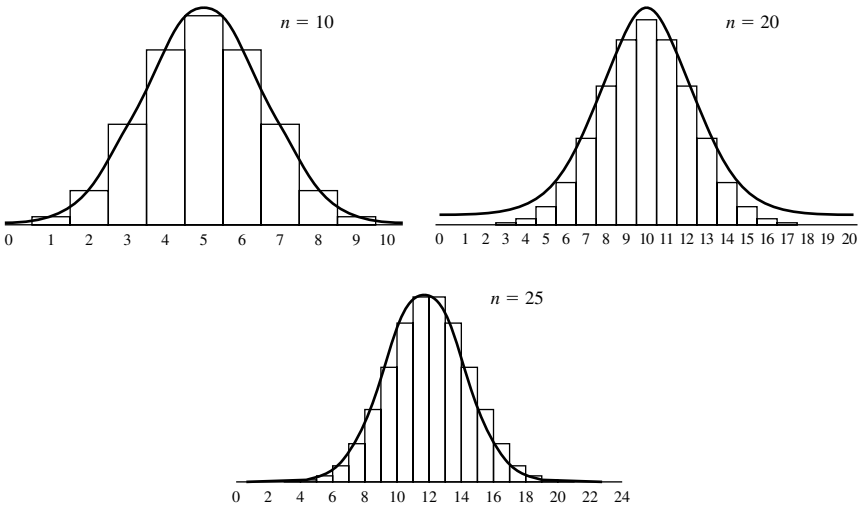
nonresponse bias– A systematic tendency for selected **elementary units** with particular characteristics not to respond in a **survey,** while other such units in the **sample,** with different characteristics, do. The units or individuals that do not respond are usually not representative of those that do. See also *nonresponse, nonresponse rate.*

nonresponse rate– The **proportion** of individuals in a **sample survey** that fail to provide the relevant information being sought by the investigator. Compare *response rate.* See also *nonresponse, nonresponse bias.*

nonsampling error– An **error** in a **sample estimate** that is not related to **sampling error.** Such errors may arise from many different sources such as flaws in the **sampling frame,** errors in the collection of **data,** mistakes in the processing of data, and so forth.

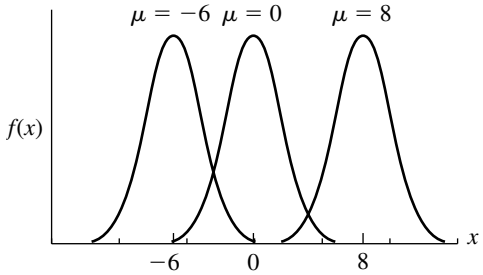
nonsense correlation– Same as *spurious correlation.*

normal approximation– A term used to denote the act of approximating a **non-normal probability distribution** by a **normal distribution;** for example, a **binomial distribution** with number of **trials** n and **probability** of a success p can be approximated by a normal distribution with **mean** np and **variance** $np(1-p)$.

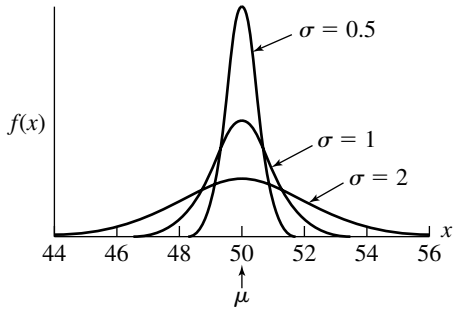


Normal approximations to binomial distribution, $n = 10, 20, 25$

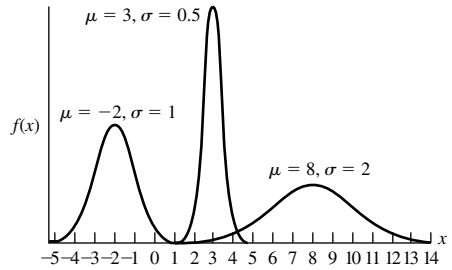
normal curve– The normal curve, or more accurately the family of normal curves, is represented by the **normal distribution.** Normal curves are **mesokurtic,** symmetrical, bell-shaped curves with tails extending indefinitely in both directions from the center, approaching but never touching the **horizontal axis.** Theoretically speaking, the curve extends from $-\infty$ to ∞ with the horizontal axis as an asymptote. The normal curve has many interesting mathematical properties and can be used to approximate the **distributions** of many other **variables.** See also *standard normal curve.*



Normal curves with $\mu = -6, 0, 8$ and $\sigma = 2$



Normal curves with $\mu = 50$ and $\sigma = 0.5, 1, 2$



Normal curves with $\mu = -2, \sigma = 1$;
$\mu = 3, \sigma = 0.5$; $\mu = 8, \sigma = 2$

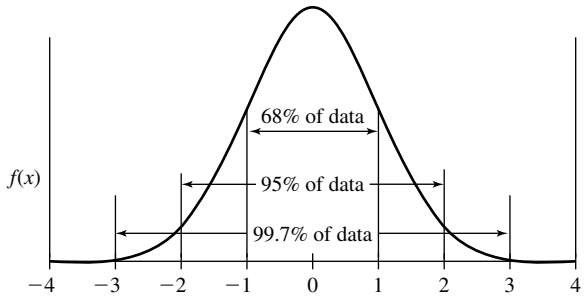
normal curve ordinate– The height of the **normal curve** at any point along its **abscissa** is the **ordinate** of the curve at that point.

normal deviate– The value of a **deviate** of the **normal distribution.**

normal distribution– A **probability distribution** of a **continuous random variable** X represented by the **probability density function**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-(x-\mu)^2/2\sigma^2\right\} \qquad -\infty < x < \infty$$

where μ and σ are, respectively, the **mean** and **standard deviation** of the **distribution.** It is also called the gaussian distribution. In any normal distribution: (1) 68% of the observations fall within σ of the mean μ, (2) 95% of the observations fall within 2σ of μ, and (3) 99.7% of the observations fall within 3σ of μ. This is known as 68–95–99.7 rule and is graphically illustrated in the figure below.



The 68–95–99.7 rule for normal distribution

normal equations– The set of simultaneous equations obtained as in the **estimation** of the **regression coefficients** by the **method of least squares.** The solution of the normal equations yields the least squares estimates of the regression coefficients.

normal equivalent deviate– See *probit transformation.*

normal form analysis– In **decision theory,** a tabular form of **preposterior analysis** that systematically calculates an expected **payoff** value for every possible strategy and then selects the strategy with the largest payoff as the optimum one.

normal interval– Same as *normal range.*

normality– A term used to denote the property of a **random variable** having a **normal distribution.**

normality assumption– Many of the parametric **tests of significance** require that the **distribution** of the **parent population**(s) involved be normal or nearly normal in **shape.**

normal law of error– Same as *law of error.*

normal limits– See *normal values.*

normal plot– Same as *normal probability plot.*

normal population– A **population** of values having a **normal distribution.**

normal probability density function– See *normal distribution.*

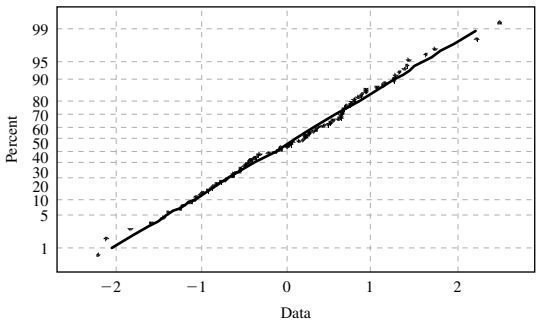
normal probability distribution– Same as *normal distribution.*

normal probability paper– Same as *arithmetic probability paper.*

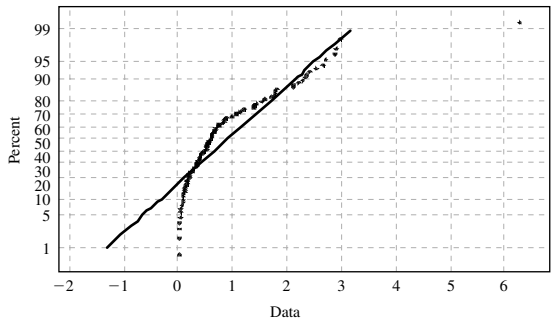
normal probability plot– A **graphical method** of assessing the assumption of **normality** of a **sample.** The ordered sample values $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ are plotted against the values $\Phi^{-1}(p_i)$ where $p_i = i - 0.5/n$ and

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt\,.$$

For a sample from a **normal distribution** the plot appears as a straight line and any departure from normality in the plot is indicative of the lack of normality of the **data.**



Normal probability plot: normally distributed data



Normal probability plot: negatively skewed data

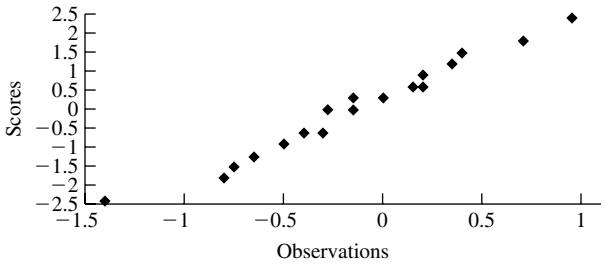
normal probability tables– Tables that give **probabilities** of a **normal distribution** for various possible combinations of values of μ (**mean**) and σ (**standard deviation**). A short version of normal probability tables for the **standard normal distribution** is given on page 180.

normal random variable– A **random variable** having a **normal distribution.**

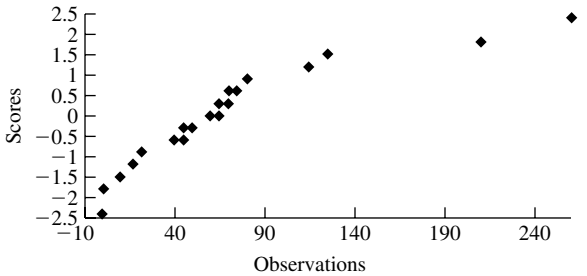
normal range– See *normal values.*

normal scores– The **expected values** of the **order statistics** $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ drawn from the **standard normal distribution.** Normal scores are used in plots to assess **normality.**

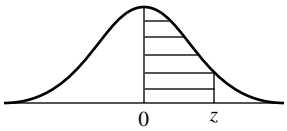
normal scores test– A **nonparametric procedure** for comparing **locations** of two **populations.** The procedure consists of first transforming the **observations** to **rank order** in the combined **sample** and then converting the **ranks** by a **transformation,** which involves **standard normal distribution.** See also *Mann–Whitney–Wilcoxon test, Wilcoxon rank-sum test.*



Normal score plot: normally distributed data



Normal score plot: exponentially distributed data



Standard normal table

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4986	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

The entries in this table are the probabilities that a random variable having the standard normal distribution assume a value between 0 and *z*

Source: Computed by using software.

normal values– Values regarded as being within the usual range of **variation** in a given **population** or population subgroup. The range of such values is called the normal range. The limits of the normal range are called normal limits. Normal values provide useful descriptive tools and, for normally distributed **data,** can be calculated by using the **sample mean** and the **standard deviation.** The normal values are often used as the basis

for evaluating the results of a diagnostic test in classifying individuals as normal or abnormal.

not statistically significant– In **hypothesis testing,** any **sample data** that do not lead to the rejection of the **null hypothesis** because it has a high **probability** of occurring when the null hypothesis is true. Compare *statistically significant.*

nuisance parameter– In **statistical estimation** and **hypothesis testing,** the term is used to designate a **parameter** that is needed to specify the **sampling distribution** of interest, but is not of direct interest for making the **inference.** The presence of a nuisance parameter makes the problem of inference more difficult and it is often necessary to find a statistical procedure that does not depend on it. For example, in testing or setting a **confidence interval** for the **mean** of a **normal population,** the unknown parent **variance** is a nuisance parameter and the problem is solved by use of the **Student's *t* distribution,** which does not depend on the parent variance.

null distribution– The **probability distribution** of a **test statistic** evaluated under the **null hypothesis.**

null hypothesis– In statistical testing, the general procedure is to assume a **hypothesis** tentatively for the purpose of rejecting or refuting it. Such a statement is called a null hypothesis. It is always a statement of some exact value or values for one or more **population parameters** usually expressed as a negative statement. This hypothesis is assumed to be true until such time as **observations** indicate that it is unlikely to be, that is, the **sample observations** show whether or not the null hypothesis should be rejected. An example is the hypothesis that a particular **treatment** has the same effect as a **placebo.** In general, the term refers to a particular hypothesis being tested, as distinct from the **alternative hypotheses** that are under consideration. The null hypothesis can be considered the hypothesis of "no difference" or, more correctly, the hypothesis that the observed difference is entirely due to **sampling error,** i.e., that it occurred purely by **chance.** In a **test of significance,** the null hypothesis is postulated to form the basis for calculating the **probability** that the difference occurred entirely by chance. When the difference is not significant, the null hypothesis is not rejected; when the difference is significant, the null hypothesis is rejected in favor of other hypotheses about the causes of the difference. Note that the null hypothesis is never proven right or wrong, or true or false, but is only rejected or not rejected at the arbitrarily chosen **level of significance,** i.e., 0.05, 0.01, 0.1, etc. It is usually denoted by H_0.

number of cases– Same as *size of a sample.*

numerical data– Data obtained by using numerical **scales of measurement.** These are data in numerical quantities involving continuous measurements or counts. See also *numerical scale.*

numerical distribution– A **frequency distribution** in which **data** are grouped according to numerical values. It is also called a quantitative distribution.

numerical observations– Same as *numerical data.*

numerical scale– It is used for characteristics that can be given numerical values and the differences between numbers have meaning. Some examples of such characteristics are

height, weight, and blood pressure level. It is also called an **interval** or **ratio scale** and is the highest **level of measurement.**

numerical taxanomy– Methods and techniques used in the numerical evaluation of the affinity or similarity between species or subspecies in biological material and the ordering and grouping of these units into taxa on the basis of their affinities. More generally, the term is used as a synonym for **cluster analysis.**

numerical variable– Same as *quantitative variable.*